Rejoinder

Blossom H. Patterson, C. Mitchell Dayton, and Barry I. Graubard

We wish to thank the discussants for their thoughtful and provocative comments. In our article we have presented a new approach to modeling dietary consumption patterns, as well as methodology permitting the application of LCA to sample survey data. We had two goals. First, extending a long-standing interest in vegetable consumption, we were interested in using LCA to find some overall (if crude) measure of the proportion of the population falling into a "regular" vegetable consumption class and the proportion falling into a less regular, or infrequent, consumption class. This type of information could be useful in formulating public health programs. A national dataset comprising 4 (independent) days of dietary intake data sampled from women age 19–50 from the CSFII offered this opportunity. Second, because these data were not from a simple random sample, we had to develop methodology to apply sample weights to the data and to estimate standard errors that take into account the complex sample design. In our "first cut" at achieving these goals, we sought to fit a simple, straightforward LC model. We recognize that there are many different and more complex approaches than our simple model, and we encourage others to pursue them. A major contribution, both in our eyes and in the eyes of the discussants, was to develop LCA methods that can be used to analyze sample survey data.

The discussions cover a broad range of topics. We begin our rejoinder by returning to our motivating problem, characterizing "usual" vegetable consumption in the United States. Central to this problem is the need to measure intake over some time period. We explain our approach to this problem, which involves a new definition of usual consumption that uses LC modeling with the consequent data reduction to binary observations. Some of the discussants questioned our data reduction and suggested alternative methods of analysis, both frequentist and Bayesian. We review and comment on some of these. The question of how (and whether) to use sample weights was of special interest to our discussants, as was the estimation of standard errors. The question of model fit arose in several of the discussions. We make the point that no adequate measure of goodness of fit for latent class models has yet been developed for sample survey data; such measures need to be developed. We address these three topics (weights, variance estimation, and goodness of fit) under the broad heading of accounting for the sample design. Finally, some of the discussants proposed new uses of our techniques, and we briefly review these.

1. CHARACTERIZING DIETARY INTAKE

As pointed out by Carriquiry and Nusser, the U. S. government relies on dietary intake data from national surveys for the development of nutritional and health policies. For example, in presenting a revised baseline for the *Healthy People 2000* objectives, Krebs-Smith et al. (1995) showed that 8.2% of the population age 20 years and older consumed less than a single serving of a vegetable per day based on 3 days of dietary intake data for 3 consecutive years. Because of the inverse association between vegetable consumption and several cancers, we were interested in using national survey data to estimate the proportion of the population that does and that does not consume vegetables on a "regular" basis, where regular can be regarded as a way of defining "usual." The National Cancer Institute is currently investigating other approaches to estimating regularity.

As Carriquiry and Nusser note, the definition of "usual" intake as the "long-run average intake of a food" is widely used, although there are other methods in the literature, some of which we cite in our article. However, there is no consensus on how to define "long run". Furthermore, average intake may not be a measure of the regularity of intake. We took a new approach to this problem, considering "regularity" of vegetable consumption to be an unobservable or latent variable. This definition is conceptual but can be operationalized via latent class modeling. We fitted a two-class model to the data. In this context, the item-conditional probabilities, measures of the probability of consuming a vegetables on each recall day given membership in a specific class, are dietary propensity scores (Sue Krebs-Smith and Kevin Dodd, personal communication). These were remarkably consistent for the class of "regular" vegetable consumers but appeared to vary for the infrequent consumers.

Concerns were raised about our dichotomization of the data, which consisted of the number of grams of each individual food consumed by each respondent. We agree that dichotomization of the data does not necessarily reduce measurement error. Carriquiry and Nusser contend that the amount of food consumed on most of the survey days, which is lost in dichotomization, may be crucial in making inferences about the impact of diet on cancer. In our method, an individual consuming small amounts of food on most of the intake days would be classified differently than an individual consuming a large amount on a single recall day, yet the average intake for both individuals could be the same. Whether frequency of consumption of vegetables or the quantity consumed is critical in disease prevention is an open question. Kant, Schatzkin, Graubard, and Schairer (2000) developed a recommended foods score (RFS) that summarizes food frequency questionnaire replies for 23 items, using the report of consumption but not the quantity consumed. They found that dietary diversity as reflected in the RFS was inversely related to cancer and other diseases as well as to all-cause mortality. Our method could be used to examine the relationship between

Journal of the American Statistical Association
September 2002, Vol. 97, No. 459, Applications and Case Studies
DOI 10.1198/016214502388618519

disease and diet. In a study similar to that described here, but with a larger sample size, respondents would be followed for morbidity or mortality, as is being done with respondents from the second National Health and Nutrition Study. A LC model could be fitted to the data, each subject assigned to a LC, and the eventual outcomes compared to these assignments. Our method could be extended to look at amounts consumed.

Elliott and Sammel suggest that a count of vegetable servings on each occasion of measurement represents a better modeling opportunity than the simpler 0–1 representation that we used. In general, we agree that this is both desirable and possible, using, for example, a Poisson representation for the counts. In the present case our judgment was that the reliability of the measurements better supported the simpler coding, but it would be interesting to compare these models.

2. MODEL CHOICE

2.1 Latent Class and Alternative Models

As discussed earlier, our choice of a two-class model was based on our interest in the proportion of the population that does and that does not consume vegetables on a regular basis. Our data constrained us to a two-class model, as noted by Carriquiry and Nusser. But this was not a problem, because the two-class model was the model of interest to us. Carriquiry and Nusser suggest an approach that would distinguish day 1 from the other recall days. In fact, a LC constrained to equate recalls for days 2–4 for each class would accomplish this objective. The similarity seen in the item-conditional probabilities for class 2 (but not class 1) suggests such a constrained model, especially for class 2. However, because this was a *post hoc* finding, we did not pursue this particular model.

Elliott and Sammel suggest extending our method to take into account all the various vegetables reported by all respondents on all 4 recall days, to create the potential for evaluating more than two classes. The resulting cross-tabulation would likely result in a very sparse table, with the accompanying problems of numerical instability and lack of convergence. An alternative method, grouping vegetables by their characteristics (e.g., deep yellow, dark-green leafy), may be a feasible extension of our model. Yet another approach would be to apply definitions of servings to grams reported by each subject, and to use mixture analysis on these variables. Measurement error in reporting portion size is a problem with this approach, as we noted in our article, and it adds a level of difficulty to the analysis.

We agree with Carriquiry and Nusser that it would be possible to fit a richer class of models using the continuous data. They summarize a method of obtaining the distribution of consumption of foods. However, their method makes the assumption that the probability of consuming a food is independent of the amount consumed, an assumption likely to be untrue for vegetables, and also requires strong distributional assumptions. Similarly, the random-effects model and the hidden Markov model suggested by Vermunt require heavy model assumptions. An assessment of the robustness of these methods to model specification is recommended.

Vermunt suggests alternative analyses that take advantage of the fact that the data were collected on six occasions. Here, as in many large surveys, the data cannot all be collected at a single time point, even when the time of interest might be as long as a season. The six data points do not represent the same time intervals during the year, and each occasion actually represents a period of several overlapping weeks (e.g., observation three for one respondent may be collected in the same month as observation four for a different respondent). Further, the two missing time points may represent occasions deleted randomly by the U.S. Department of Agriculture for subjects with five or six responses or actually may be missing data, so that the mechanism of missingness differs between respondents and is not known to the analyst. For this reason, we chose not to analyze data for all six occasions and cannot agree with Vermunt's interpretation that in these data, "consumption of vegetables... depends on the time of year." Pairwise z tests using jackknifed standard errors (our Table 2) among the four conditional probabilities within each LC result in no absolute z value greater than 1.13, suggesting, on a *post hoc* basis, that homogeneity is not an unreasonable assumption for the rates of vegetable consumption.

Elliott and Sammel propose a *post hoc* Bayesian approach, using Bayes's theorem to calculate a predicted LC membership for each sample member. Given these classifications, odds ratios can be computed for outside variables such as age and region. Because of concerns about the validity of the two-stage procedure, we did not report analyses of this type. However, a recently completed simulation study (Kuo 2001) suggests that, at least for simple random samples, the two-stage procedure for logistic covariate models performs quite well in estimating the parameters for the covariate function for cases with well-defined latent structures (i.e., cases where the conditional probabilities for the two classes are distinctly different). The vegetable data seem to satisfy this requirement.

From a theoretical perspective, the best strategy would be to use the outside variables as covariates directly within the latent class model (Dayton and Macready 1988). In brief, for a two-class model, the latent class proportion for class 1, say, is modeled by a function of the form

$$\pi^X_{1|Z} = g(Z, \beta),$$

where Z is in general a vector-valued covariate, β is a vector of parameters, and $g(\cdot)$ is a monotone function with a 0, 1 range over the domain of Z. For example, a logistic covariate model with J covariates could be defined as

$$\pi^X_{1|Z} = 1/[1 + e^{-\beta_0 - \Sigma^J_{j=1} \beta_j Z_j}],$$

where $\pi^X_{1|Z}$ is the proportion of cases in the first latent class (X) conditional on the covariate vector, Z.

Conditional probabilities for the manifest variables and the parameters of the covariate function are estimated simultaneously. Programs such as Latent Gold (Vermunt and Magidson 2000) and LEM (Vermunt 1997) provide estimates for logistic covariate models with case weights but do not take into account clustering. In the context of a complex survey design, one is faced with assessing the contribution of a covariate. The jackknife is recommended as an easily applicable and valid method for generating standard errors of the regression coefficients of the covariates for complex samples.

2.2 Goodness of Fit

Model fit is mentioned in several of the discussions. As we state in our article, we are unaware of any good measure of fit that is appropriate for LCA of complex sample survey data. We also note that for data from a simple random sample, the likelihood ratio statistic cannot be used for comparison of models with differing numbers of classes. Vermunt fits a weighted model based on a Poisson sampling model and concludes that a two-class model "does not fit the data." Although his analysis takes into account sampling weights, it ignores the role of stratification and clustering of the sample selection in survey design. Our analysis, based on a Wald test that took into account both stratification and clustering, suggested satisfactory fit for a two-class model.

3. ACCOUNTING FOR THE SAMPLE DESIGN

3.1 Sample Weights

The CSFII has a complex sample design involving stratified multistage cluster sampling with sample weighting for nonresponse and postratification adjustment. Vermunt argues that weighting should be used for estimation of the LC proportions but not for estimation of the item-conditional probabilities. He prefers a two-stage approach, in which the unweighted data are used to estimate the conditional probabilities and these are then held constant during a weighted analysis that estimates the LC proportions. He argues that if the sample weights are informative for estimating the items conditional probabilities, then these probabilities are not homogenous across subgroups of the population, and the LC model is misspecified for the population. We address these issues in Section 6 of our article and also address the bias and efficiency trade-off between weighted and unweighted analyses. Vermunt also proposes using a method described by Clogg and Eliason (1987); however, this method does not adjust for clustering in the data and also assumes that the model is correctly specified. It is important to note that in general, it is not possible to know whether a model is "correctly" specified, and even if this were possible, the "correct" model would likely be unduly complex and difficult to interpret. When the posited LC model is misspecified, Vermunt's two-stage approach does not estimate the "census" model, that is, the model that would have been obtained if the entire population had been sampled. In contrast, the weighted pseudolikelihood approach that we use does estimate the census model. This approach has the advantage that if the model is misspecified, estimates from different probability sample designs on average will be approximately the same. Vermunt's suggestion of dealing with heterogeneity of the item-conditional probabilities by identifying homogeneous groups and then using multiple-group LC analysis seems impractical and difficult to carry out.

As shown in Figure 1, the impact of weighting is to lower the magnitude of the conditional probabilities, although the effect is much greater in the low-consumption class than in the high-consumption class.

Elliott and Sammel report a more elaborate analysis based on stratification of the sample by the magnitude of the weights themselves. This appears to show that the estimated item-conditional probabilities differ between the low weight stratum and the medium and high weight strata. They did not test whether these differences are statistically significant. Based on our null results for a Wald test comparing weighted to unweighted estimates, we doubt that theirs would be statistically significant. As pointed out in our article, the Wald test for informativeness of the weights has low power. Based on our results and those of Elliott and Sammel, we are inclined to believe that the weighted estimate is the more reasonable estimate for the population.

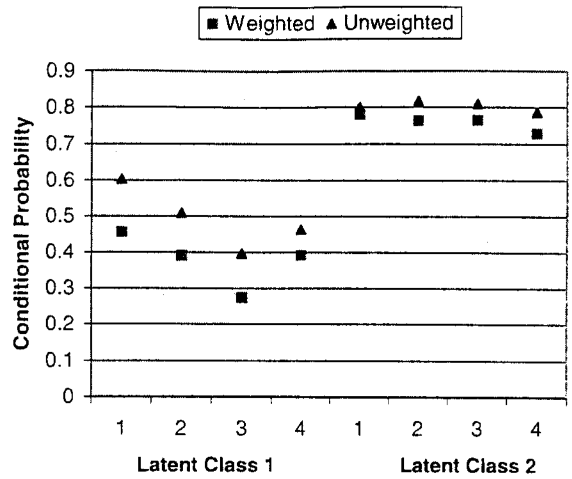

Figure 1. Impact of Sampling Weights on Conditional Probabilities.

Elliott and Sammel also propose an interesting alternative to the "all-or-nothing" approach to weighting. They divide the data into design strata and use an estimator that is a combination of weighted and unweighted estimates. The weighted estimate has more influence on the overall estimates if there is evidence of substantial variability across the strata in the parameter of interest and the unweighted estimate has less influence if there is little evidence of variability. Because this approach requires a prior distribution over the strata-specific parameters, its robustness to the distribution of the assumed prior should be investigated before using it. Elliott and Sammel also propose an extension to this model, a hierarchical model that requires both hyperpriors and priors. Such a model would require substantial robustness testing.

3.2 Variance Estimation

Vermunt recommends using linearization variance estimation rather than the jackknife variance estimation that we used. We agree that linearization variances will be faster to compute and can be programmed for various LCAs. We chose to use a jackknife method because of its ease of use; that is, it does not requiring extensive programming. In addition, jackknife variance estimation, through the use of jackknife replicate weights (Rust and Rao 1996), is more flexible than linearization in that it is able to account for variation inherent in commonly used adjustments to the sample weights, such as nonresponse adjustments and postratification. Similar types of replicate weights can be formed from other variance replication methods, such as balanced half-sample replication. National surveys such as the third National Health and Nutrition Survey (Ezzati, Massey, Waksberg, Chu, and Maurer 1995) are now routinely providing replicate weights for variance estimation.

4. ALTERNATIVE APPLICATIONS

Seastrom suggests other applications of our LC model. Especially useful is her idea of modeling LCs that reflect level of risk for an adverse health, behavioral, or social outcome. It is reasonable to hypothesize that a population may have risk patterns that can be classified into discrete unobservable categories. Also, by identifying these LCs and their relative sizes in the population, intervention programs can be constructed that could be directed at the highest risk classes of nontrivial size. Because complex surveys are used extensively in behavioral and social research, our results for using design-based analyses are potentially of great value for carrying out such analyses with survey data.

ADDITIONAL REFERENCES

Dayton, C. M., and Macready, G. B. (1988), "Concomitant-Variable Latent Class Models.," *Journal of the American Statistical Association*, 83, 173–178.

Ezzati, T. M., Massey, J. T., Waksberg, J., Chu, A., and Maurer, K. R. (1992), "Sample Design: Third National Health and Nutrition Examination Survey," *Vital and Health Statistics* 2(113).

Kant, A. K., Schatzkin, A., Graubard, B. I., and Schairer, C. (2000), "A Prospective Study of Diet Quality and Mortality in Women," *Journal of the American Medical Association*, 283, 2109–2115.

Kuo, W.-L. (2002), "Two-stage Concomitant Variable Latent Class Analysis," PhD dissertation, University of Maryland, College Park, MD.

Rust, K. F., and Rao, J. N. K. (1996), "Variance Estimation for Complex Surveys Using Replication Techniques," *Statistical Methods in Medical Research*, 5, 283–310.